CLAIMS

1. In a cluster of computing nodes having shared access to one or more volumes of data storage using a parallel file system, a method for managing the data storage, comprising:

selecting a first one of the nodes to serve as a session manager node;

selecting a second one of the nodes to serve as a session node for a data management application;

creating a session of the data management application on the session node by sending a message from the session node to the session manager node, causing the session manager node to distribute information regarding the session among the nodes in the cluster; and

responsive to the information distributed by the session manager node, receiving events at the session node from the nodes in the cluster for processing by the data management application.

2. A method according to claim 1, and comprising storing the information regarding the session at the session manager node.

3. A method according to claim 2, wherein the information regarding the session is stored at both the session node and at the session manager node, and comprising, following a failure at the session node, receiving the stored information from the session manager node in order to recover the session.

4. A method according to claim 3, wherein at least a portion of the information regarding the session is stored at one or more of the nodes in the cluster other

than at the session node and the session manager node, and comprising, following a failure at the first one of the nodes, selecting a third one of the nodes to serve as the session manager node, and collecting the information from at least one of the session node and the other nodes in the cluster at which the information is stored for use by the third one of the nodes in serving as the session manager node.

5. A method according to claim 1, wherein creating the session comprises creating the session in accordance with a data management application programming interface (DMAPI) of the parallel file system, and wherein sending the message comprises invoking a session management function of the DMAPI on the session manager node.

6. A method according to claim 1, wherein the information comprises a list of the events in each file system of relevance to the data management application and respective dispositions of the events on the list, and wherein receiving the events at the session node comprises receiving messages reporting the events appearing on the list responsive to the dispositions.

7. A method according to claim 6, and comprising receiving a data management application programming interface (DMAPI) function call from one or more of the nodes other than the session node setting one or more of the dispositions.

8. A method according to claim 6, wherein the session is one of a plurality of sessions in the cluster, and wherein the session manager node coordinates a consistent partitioning of the dispositions among the sessions.

9. A method according to claim 1, wherein selecting the second one of the nodes comprises selecting a plurality of the nodes to serve as respective session nodes in a plurality of data management sessions, and wherein creating the session comprises informing the session manager node of the plurality of the sessions, causing the session manager node to distribute the information regarding the plurality of the sessions.

10. A method according to claim 9, wherein the first one of the nodes serves as one of the session nodes, in addition to serving as the session manager node.

11. A method according to claim 9, wherein selecting the plurality of the nodes comprises selecting multiple session nodes for a distributed data management application running in the cluster.

12. A method according to claim 9, wherein selecting the plurality of the nodes comprises selecting the second one of the nodes to serve as the respective session node for a first data management application, and selecting a third one of the nodes to serve as the respective session node for a second data management application.

13. A method according to claim 1, and comprising modifying the session by sending a further message from the session node to the session manager node, causing the session manager node to distribute a notification regarding the modified session to the nodes in the cluster.

14. A method according to claim 1, and comprising destroying the session by sending a further message from the session node to the session manager node, causing the

session manager node to distribute a notification regarding the destroyed session to the nodes in the cluster.

15. A method according to claim 1, wherein creating the session of the data management application comprises initiating a data migration application, so as to free storage space on at least one of the volumes of data storage.

16. A method according to claim 1, wherein the information comprises a session identifier, generated at the session manager node, which is unique within the cluster.

17. In a cluster of a plurality of computing nodes having shared access to one or more volumes of data storage using a parallel file system, a method for managing the data storage, comprising:

initiating sessions of a parallel data management application on the plurality of the nodes, so that an instance of the data management application runs on each of the nodes;

generating a data management event responsive to a request submitted to the parallel file system on at least one of the nodes to perform a file operation on a file in one of the volumes of data storage;

handling the event by means of the instance of the data management application running on the at least one of the nodes.

18. A method according to claim 17, and comprising sending an event message from the at least one of the nodes to the other nodes, so as to inform the data

management application sessions on the other nodes of the event.

19. A method according to claim 17, wherein generating the data management event comprises running a user application on the at least one of the nodes, and receiving the request from the user application.

20. Computing apparatus, comprising:

one or more volumes of data storage, arranged to store data; and

a plurality of computing nodes, linked to access the volumes of data storage using a parallel file system, and arranged so as to select a first one of the nodes to serve as a session manager node and to select a second one of the nodes to serve as a session node for a data management application, so that a session of the data management application is created on the session node by sending a message from the session node to the session manager node, causing the session manager node to distribute information regarding the session among the nodes in the cluster, responsive to which the session node receives events from the nodes in the cluster for processing by the data management application.

21. Apparatus according to claim 20, wherein the session manager node is arranged to store the information regarding the session.

22. Apparatus according to claim 21, wherein the information regarding the session is stored at both the session node and at the session manager node, and wherein following a failure at the session node, the stored

information is received from the session manager node in order to recover the session.

23. Apparatus according to claim 22, wherein at least a portion of the information regarding the session is stored at one or more of the nodes in the cluster other than at the session node and the session manager node, and wherein following a failure at the first one of the nodes, the nodes are arranged to select a third one of the nodes to serve as the session manager node, and to collect the information from at least one of the session nodes and the other nodes in the cluster at which the information is stored for use by the third one of the nodes in serving as the session manager node.

24. Apparatus according to claim 20, wherein the session is created in accordance with a data management application programming interface (DMAPI) of the parallel file system, and wherein sending the message invokes a session management function of the DMAPI on the session manager node.

25. Apparatus according to claim 20, wherein the information comprises a list of the events in each file system of relevance to the data management application and respective dispositions of the events on the list, and wherein the nodes are arranged to report to the session node the events appearing on the list responsive to the dispositions.

26. Apparatus according to claim 25, wherein the dispositions can be set by any of the nodes.

27. Apparatus according to claim 25, wherein the session is one of a plurality of sessions in the cluster, and

wherein the session manager node coordinates a consistent partitioning of the dispositions among the sessions.

28. Apparatus according to claim 20, wherein the nodes are arranged so that a plurality of the nodes can be selected to serve as respective session nodes in a plurality of data management sessions, and wherein the session manager node is informed of the plurality of the sessions, causing the session manager node to distribute the information regarding the plurality of the sessions.

29. Apparatus according to claim 28, wherein the first one of the nodes is arranged to serve as one of the session nodes, in addition to serving as the session manager node.

30. Apparatus according to claim 28, wherein the plurality of the nodes comprises multiple session nodes selected for a distributed data management application running in the cluster.

31. Apparatus according to claim 28, wherein the plurality of the nodes comprises the second one of the nodes, selected to serve as the respective session node for a first data management application, and a third one of the nodes selected to serve as the respective session node for a second data management application.

32. Apparatus according to claim 20, wherein the session node is arranged to modify the session by sending a further message to the session manager node, causing the session manager node to distribute a notification regarding the modified session to the nodes in the cluster.

33. Apparatus according to claim 20, wherein the session node is arranged to destroy the session by sending a further message to the session manager node, causing the session manager node to distribute a notification regarding the destroyed session to the nodes in the cluster.

34. Apparatus according to claim 20, wherein the data management application comprises a data migration application, for freeing storage space on at least one of the volumes of data storage.

35. Apparatus according to claim 20, wherein the information comprises a session identifier, generated at the session manager node, which is unique within the cluster.

36. Computing apparatus, comprising:

one or more volumes of data storage, arranged to store data; and

a plurality of computing nodes, linked to access the volumes of data storage using a parallel file system, and arranged to initiate sessions of a parallel data management application on the plurality of the nodes, so that an instance of the data management application runs on each of the nodes, and a data management event is generated responsive to a request submitted to the parallel file system on at least one of the nodes to perform a file operation on a file in one of the volumes of data storage, causing the event to be handled by the instance of the data management application running on the at least one of the nodes.

37. Apparatus according to claim 36, wherein the at least one of the nodes is arranged to send an event message to the other nodes, so as to inform the data management application sessions on the other nodes of the event.

38. Apparatus according to claim 36, wherein the data management event is generated by a user application running on the at least one of the nodes, which submits the request.

39. A computer software product for use in a cluster of computing nodes having shared access to one or more volumes of data storage using a parallel file system, the product comprising a computer-readable medium in which program instructions are stored, which instructions, when read by the computing nodes, cause a first one of the nodes to be selected to serve as a session manager node and a second one of the nodes to be selected to serve as a session node for a data management application, and cause a session of the data management application to be created on the session node by sending a message from the session node to the session manager node, causing the session manager node to distribute information regarding the session among the nodes in the cluster, responsive to which the session node receives events from the nodes in the cluster for processing by the data management application.

40. A product according to claim 39, wherein the instructions cause the session manager node to store the information regarding the session.

41. A product according to claim 40, wherein the instructions cause the information regarding the session to be stored at both the session node and at the session manager node, and wherein following a failure at the session node, the instructions cause the stored information to be received from the session manager node in order to recover the session.

42. A product according to claim 41, wherein the instructions cause at least a portion of the information regarding the session to be stored at one or more of the nodes in the cluster other than at the session node and the session manager node, and wherein following a failure at the first one of the nodes, the instructions cause the nodes to select a third one of the nodes to serve as the session manager node, and to collect the information from at least one of the session node and the other nodes in the cluster at which the information is stored for use by the third one of the nodes in serving as the session manager node.

43. A product according to claim 39, wherein the product comprises a data management application programming interface (DMAPI) of the parallel file system, and wherein the instructions cause the session node to send the message by invoking a session management function of the DMAPI on the session manager node.

44. A product according to claim 39, wherein the information comprises a list of the events in each file system of relevance to the data management application and respective dispositions of the events on the list, and wherein the instructions cause the nodes to report to

the session node the events appearing on the list responsive to the dispositions.

45. A product according to claim 44, wherein the dispositions can be set by any of the nodes.

46. A product according to claim 44, wherein the session is one of a plurality of sessions in the cluster, and wherein the session manager node coordinates a consistent partitioning of the dispositions among the sessions.

47. A product according to claim 39, wherein the instructions are such as to cause a plurality of the nodes to be selected to serve as respective session nodes in a plurality of data management sessions, and to cause the session nodes to inform the session manager node of the plurality of the sessions, causing the session manager node to distribute the information regarding the plurality of the sessions.

48. A product according to claim 47, wherein the instructions allow the first one of the nodes to serve as one of the session nodes, in addition to serving as the session manager node.

49. A product according to claim 47, wherein the plurality of the nodes comprises multiple session nodes selected for a distributed data management application running in the cluster.

50. A product according to claim 47, wherein the plurality of the nodes comprises the second one of the nodes, selected to serve as the respective session node for a first data management application, and a third one of the nodes selected to serve as the respective session node for a second data management application.

51. A product according to claim 39, wherein the instructions cause the session node to modify the session by sending a further message to the session manager node, causing the session manager node to distribute a notification regarding the modified session to the nodes in the cluster.

52. A product according to claim 39, wherein the instructions cause the session node to destroy the session by sending a further message to the session manager node, causing the session manager node to distribute a notification regarding the destroyed session to the nodes in the cluster.

53. A product according to claim 39, wherein the data management application comprises a data migration application, for freeing storage space on at least one of the volumes of data storage.

54. Apparatus according to claim 39, wherein the information comprises a session identifier, generated at the session manager node, which is unique within the cluster.

55. A computer software product for use in a cluster of computing nodes having shared access to one or more volumes of data storage using a parallel file system, the product comprising a computer-readable medium in which program instructions are stored, which instructions, when read by the computing nodes, cause sessions of a parallel data management application to be initiated on the plurality of the nodes, so that an instance of the data management application runs on each of the nodes, and a data management event is generated responsive to a

request submitted to the parallel file system on at least one of the nodes to perform a file operation on a file in one of the volumes of data storage, causing the event to be handled by the instance of the data management application running on the at least one of the nodes.

56. A product according to claim 55, wherein the instructions cause the at least one of the nodes to send an event message to the other nodes, so as to inform the data management application sessions on the other nodes of the event.

57. Apparatus according to claim 55, wherein the data management event is generated by a user application running on the at least one of the nodes, which submits the request.